



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2011

Post-transcriptional repair of a split heat shock protein 90 gene by mRNA trans-splicing

Nageshan, R K ; Roy, N ; Hehl, A B ; Tatu, U

Abstract: Heat shock protein 90 participates in diverse biological processes ranging from protein folding, cell cycle, signal transduction and development to evolution in all eukaryotes. It is also critically involved in regulating growth of protozoa such as *Dictyostelium discoideum*, *Leishmania donovani*, *Plasmodium falciparum*, *Trypanosoma cruzi*, and *Trypanosoma evansi*. Selective inhibition of Hsp90 has also been explored as an intervention strategy against important human diseases such as cancer, malaria, or trypanosomiasis. *Giardia lamblia*, a simple protozoan parasite of humans and animals, is an important cause of diarrheal disease with significant morbidity and some mortality in tropical countries. Here we show that the *G. lamblia* cytosolic hsp90 (glhsp90) is split in two similar sized fragments located 777 kb apart on the same scaffold. Intrigued by this unique arrangement, which appears to be specific for the *Giardiinae*, we have investigated the biosynthesis of GlHsp90. We used genome sequencing to confirm the split nature of the giardial hsp90. However, a specific antibody raised against the peptide detected a product with a mass of about 80 kDa, suggesting a post-transcriptional rescue of the genomic defect. We show evidence for the joining of the two independent Hsp90 transcripts in-trans to one long mature mRNA presumably by RNA splicing. The splicing junction carries hallmarks of classical cis-spliced introns, suggesting that the regular cis-splicing machinery may be sufficient for repair of the open reading frame. A complementary 26-nt sequence in the "intron" regions adjacent to the splice sites may assist in positioning the two pre-mRNAs for processing. This is the first example of post-transcriptional rescue of a split gene by trans-splicing.

DOI: <https://doi.org/10.1074/jbc.C110.208389>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-58785>

Journal Article

Accepted Version

Originally published at:

Nageshan, R K; Roy, N; Hehl, A B; Tatu, U (2011). Post-transcriptional repair of a split heat shock protein 90 gene by mRNA trans-splicing. *Journal of Biological Chemistry*, 286(9):7116-7122.

DOI: <https://doi.org/10.1074/jbc.C110.208389>

Post-transcriptional repair of a split heat shock protein 90 gene by mRNA *trans*-splicing

Rishi Kumar Nageshan¹, Nainita Roy¹, Adrian B. Hehl^{2*} and Utpal Tatu^{1*}

Department of Biochemistry¹, Indian Institute of Science, Bangalore: 560012, India
Institute of Parasitology², University of Zürich, Winterthurerstrasse 266a, CH-8046 Zürich

Running title: A *trans*-spliced heat shock protein 90 gene product in *G. lamblia*

Address correspondence to : Utpal Tatu^{1*}, Department of Biochemistry, Indian Institute of Science, Bangalore: 560012, India. E mail: tatu@biochem.iisc.ernet.in Tel. office/lab: 011 91 80 22932823 and Adrian B. Hehl^{2*} Institute of Parasitology, University of Zürich, Winterthurerstrasse 266a, CH-8046 Zürich. E mail: adrian.hehl@access.uzh.ch Tel. office/lab: +41 44 635 8526/30 Tel secretariat: +41 44 635 8501

Heat shock protein 90 participates in diverse biological processes ranging from protein folding, cell cycle, signal transduction, development to evolution in all eukaryotes. It is also critically involved in regulating growth of protozoa such as *Dictyostelium discoideum*, *Leishmania donovani*, *Plasmodium falciparum*, *Trypanosoma cruzi* and *Trypanosoma evansi*. Selective inhibition of Hsp90 has also been explored as an intervention strategy against important human diseases such as cancer, malaria or trypanosomiasis. *Giardia lamblia*, a simple protozoan parasite of humans and animals is an important cause of diarrheal disease with significant morbidity and some mortality in tropical countries. Here we show that the *G. lamblia* cytosolic Hsp90 (GIHsp90) is split in two similar sized fragments located 777 kb apart on the same scaffold. Intrigued by this unique arrangement which appears to be specific for the *Giardiinae*, we have investigated the biosynthesis of GIHsp90. We used genome sequencing to confirm the split nature of the giardial Hsp90. However, a specific antibody raised against the peptide detected a single product with a mass of about 80 kDa, suggesting a post transcriptional rescue of the genomic defect. We show evidence for the joining of the two independent Hsp90

transcripts in *trans* to one long mature mRNA presumably by RNA splicing. The splicing junction carries hallmarks of classical *cis* spliced introns suggesting that the regular *cis* splicing machinery may be sufficient for the repair of the open reading frame. A complementary 26 nt sequence in the “intron” regions adjacent to the splice sites may assist in positioning the two pre mRNAs for processing. This is the first example of post transcriptional rescue of a split gene by *trans*-splicing.

INTRODUCTION

Whole genome sequencing has galvanized investigation of biological processes and host pathogen interactions. Protozoan parasites cause life threatening diseases in humans and animals accounting for hundreds of million deaths annually worldwide. In the last 10 years the genome sequences of many important pathogens such as *Giardia lamblia*, *Leishmania donovani*, *Plasmodium falciparum*, *Trichomonas vaginalis*, and *Trypanosoma cruzi* have been completed (1–5) and have helped biological research towards development of new drug candidates. The first partial genome sequence of *G. lamblia* was reported in 2000 (4), and more recently, in 2007, the full genome sequence was published and is available on GiardiaDB

(<http://giardiadb.org/giardiadb/>) (5). *G.*

lamblia (syn *G. intestinalis*, *G. duodenalis*) is an extracellular enteroparasite causing giardiasis, a diarrheal disease, in humans and animals. Initially thought to be a primitive eukaryote, *G. lamblia* turned out to be a highly simplified extracellular parasite which has undergone significant reductive evolution (6). The rapidly accumulating genomic and postgenomic data on *Giardia* and related genera now provides the necessary background to investigate complex biological processes in a comparative manner across species boundaries.

Heat shock protein 90 (Hsp90) is an essential molecular chaperone which is implicated in growth and development of many protozoan species such as *Dictyostelium*, *Leishmania*, *Plasmodium* and *Trypanosoma* (7-10). Hsp90 has also been explored as a drug target in *Plasmodium* and *Trypanosoma* (10). Recent years have seen many new roles being ascribed to Hsp90. In addition to its ability to fold proteins, regulate cell cycle and development, recent studies also suggests an ability to guide evolution in eukaryotes (11-13). The function of Hsp90 as a sensor of environmental cues is especially crucial in protozoan parasites which often need to respond to radical changes of milieu within and outside their hosts (7, 10).

In all organisms investigated to date, Hsp90 proteins are encoded by a single open reading frame (ORF) which usually contains multiple introns. In the genome sequence of three *Giardia* isolates, no contiguous Hsp90 ORF was predicted but two fragments separated by a large stretch of sequence on the same scaffold were detected and annotated as Hsp90. Here we investigated the split nature of GIHsp90 and the consequences of this unique genetic rearrangement. We report a post-transcriptional repair mechanism which generates a *trans*-spliced, mature mRNA

coding for a bona fide Hsp90 protein from the two Hsp90 pre mRNAs.

EXPERIMENTAL PROCEDURES

Cultivation of parasites - *G. lamblia* Portland P1 or WB-C6 (assemblage A) parasites were cultured in TYI – S33 (14) supplemented with 12% fetal bovine serum and sub-cultured with 5×10^4 cells per tube from log phase parasites. The parasites were harvested by chilling on ice for 20 min. followed by repeatedly inverting the tubes to dislodge the parasites and finally pelleted down at 700 X g for 5mins.

PCR and Whole genome sequencing - Genomic DNA was isolated as described previously (15) with minor modifications. To confirm the position of HspN following primers against flanking (200bp upstream and downstream) region was used sense 5'-GGGCATGAGCTGCCGG-3' and antisense primer 5'-GTCTGCACGCCATAGACGC-3'.

Similarly HspC (200bp upstream and downstream) was gene position was also confirmed from sense primer 5'-CCGCATGCTGAGGGTGC-3' and antisense primer 5'-CCGTGCAGCTCTAGCACAATTAC-3'.

Total RNA was prepared by TRI-Reagent (Ambion) according to the manufacturer's protocol. Five µg of total RNA was used for cDNA preparation using Oligo dT primers (Fermentas cDNA kit). A full length Hsp90 ORF was amplified with specific primers overlapping the Hsp90 start codon of HspN (ORF 98054 in GiardiaDB), sense 5'-ATGCCCCGCTGAAGTCTTCGAGTTCCAG-3' and HspC (ORF 13864 in GiardiaDB), antisense 5'-TCAGTCAACTTCGTCAACGTCCTCC TC-3'. As an independent determination of the exact site of the transition from the transcript derived from ORF 98054 into that derived from ORF 13864, a PCR fragment from a cDNA template was generated with internal primers HspC-int,

sense 5'-GCGAATTCAGGTCCACGAGCACGTG AAC-3' and HspN-int, antisense 5'-GCGAATTCCTGTGATGTAGTAGATC GAC-3'. The resulting 640 bp product was cloned into the EcoRI restriction sites of pBluescript (Statagene) and sequenced. To rule out the possibility of *cis*-splicing, PCR was carried out using sense and antisense primers, 5'-ATGCTCCAGAAGAATCGC-3' and 5'-CCGTGCAGCTCTAGCACAATTAC-3' respectively. RNA was extracted using TRI reagent as described previously. RT PCR was carried out using the above primers targeted towards amplifying a specific product to look for the presence of any putative long mRNA transcript.

Whole genome sequencing was carried out with the Paired End genomic sequencing methodology with 72bp read lengths in the Illumina GA II X sequencer. More than 17 million reads were obtained and 13 million high quality reads were passed through the pipeline for alignment with the reference genome sequence (*G. lamblia* WB, ATCC 50803). The genome coverage was calculated at 165 X and SNPs and InDels were tabulated.

Western blot analysis - Parasites were lysed with 20 mM Tris HCl pH 6.8 with 1% Triton X 100 and protease inhibitor cocktail (G-biosciences). A high speed supernatant was resolved on a reducing 10% SDS PAGE gel and blotted to nitrocellulose filters. A rabbit anti GIHsp90 antibody was raised against a peptide, NKQPALWTRDPKDVTEDE, specific (Custom synthesis, Mumbai) to HspN and was used to probe the filters.

In-gel digestion - A narrow slice corresponding to a GIHsp90 band was cut from the stained SDS-PAGE gel and further sliced into smaller gel plugs. After several washes with 100 mM ammonium bicarbonate (NH_4HCO_3) (Sigma- Aldrich) buffer in 50 % Acetonitrile (ACN), the gel plugs were subjected to a reduction step using 10 mM dithiothreitol (DTT) (Sigma-

Aldrich) in 100 mM NH_4HCO_3 buffer (45 min at 56°C). Alkylation was performed with a solution of 55 mM iodoacetamide (IAA) (Sigma-Aldrich) in 100 mM NH_4HCO_3 (30 min at room temperature in the dark) followed by in-gel digestion with 20 μl trypsin (10 ng/ μl) (Promega) in 50 mM NH_4HCO_3 (overnight at 37°C). The reaction was stopped by storing at -20°C and peptides were extracted in 5 % formic acid. Samples were vacuum dried and reconstituted in 5 % formic acid.

Mass spectrometry and database searching - The protein digest was analyzed by automated nanoflow LC-MS/MS. The sample was loaded onto PepMap C18 reverse phase column connected to Tempo nano HPLC system. The peptides were eluted from the analytical column by a linear gradient of 95% Solvent A (98% H_2O , 2% ACN and 0.1% formic acid) to 10% Solvent A and Solvent B (2% H_2O , 98% ACN, 0.1% Formic acid). The spectra were acquired on a Q-STAR Elite mass spectrometer equipped with Applied Biosystems NanoSpray II ion source. The data was acquired in a data dependent mode, one MS spectrum followed by three MS/MS spectra. Data analysis was performed in Analyst QS 2.0 software. For identification of proteins the processed data was searched against NCBI nr database using the Mascot protein identification tool version 2.0 with precursor and fragment mass tolerances of 0.15 Da, cysteine carbamidomethylation as fixed modification and methionine oxidation, lysine acetylation, glutamine and asparagine deamidation as variable modifications. The resulting MS/MS based peptide identifications were manually verified in ProteinPilot 2.0 Software.

Determination of effect of Hsp90 inhibition - Parasites from log phase were harvested and 1.5×10^4 cells were treated with 0 to 10mM concentration of 17-(allylamino)-17-demethoxygeldanamycin (17AAG) with 10 fold dilutions. 0.5% DMSO was

used as control in 300 μ L of TYI - S33 culture medium. The assay was performed in 96 well plates which were sealed with parafilm and incubated in a humidified 5% CO₂ incubator. The parasite viability was scored by trypan blue exclusion and by assessing the motility of the parasites as described previously (16).

RESULTS

A genome search reveals two partial ORFs homologous to Hsp90 - The *Giardia* genome (isolate WB, assemblage A) codes for two ORFs coding for predicted proteins with strong homologies to Hsp90. The amino terminal domain with a part of a middle domain, annotated in GiardiaDB designated here as HspN (GL5083_98054, 376 amino acids) and the remaining middle domain with the carboxy terminal domain ending with DEVD, represented here as HspC (ORF GL50803_13864, 324 amino acids) (5). The predicted ORFs as well as the presence of a large intervening genomic region of 777 kb suggested that the GlHsp90 fragments were transcribed as separate mRNAs (Figure 1A). We also noticed that a catalytic Arg residue which is a hallmark of all Hsp90s reported so far was not encoded in either of the predicted ORFs in GiardiaDB.

Genome re-sequencing using a Portland isolate of *G. lamblia* ruled out the possibility that the split gene was erroneously annotated. Paired end whole genome sequencing was carried out as described in experimental procedures. The tagged genomic library was sequenced at 72 bp read lengths. About 15.7 million high quality reads were obtained out of which 15.3 million reads aligned to the previously published reference genome of *G. lamblia* which covered about 97% of the whole genome. We obtained significant coverage of about 168X of the reference sequence. As shown in Figure 1A, we were able to confirm Hsp90 represented as two fragments. We did,

however, observe several single nucleotide polymorphisms and other minor InDels (Supplementary Table 1). Recently, the sequences of two additional assemblages (B and E) were integrated into GiardiaDB. The gene pages for the two predicted Hsp90 ORFs show a high degree of synteny in the surrounding 10 kbp. Further the genomic organization was confirmed by PCR using primers specific to flanking region of both HspN and HspC ORFs. We used primers 200 bp upstream and downstream of these ORFs. PCR was carried out using genomic DNA isolated from *Giardia lamblia* culture as described in experimental procedures. As shown in the Figure 1C, specific amplicons of expected size of 1.5 kb and 1.4kb were obtained on using specific primers for HspN and HspC respectively. These results confirmed the split gene organization of Hsp90 in the genome of *G. lamblia*. Altogether, the whole genome sequencing as well as PCR approach confirmed the presence of one HspN and HspC fragment each, separated by a stretch of 777 kb of genomic sequence.

Identification of a large Hsp90 protein in Giardia - In contradiction to the genomic organization of Hsp90 as two fragments we identified a large protein using a polyclonal anti-GlHsp90 antibody raised against a peptide in the HspN sequence. Western blot analysis of separated *G. lamblia* lysates incubated with the antibodies revealed a specific band at ~80 kDa (Figure 2A). To confirm the presence of an Hsp90 protein of this mass we excised a narrow Coomassie stained gel strip for in gel digestion with trypsin and analysis by mass spectrometry. The analysis showed significant presence of tryptic peptides corresponding to both HspN and HspC derived proteins (Figure 2B, E). Indirect Immunofluorescence analysis using the anti Hsp90 serum showed a cytoplasmic localization of the detected protein which appeared to be excluded from the nuclei (Figure 2D). This

is consistent with the subcellular distribution of Hsp90 proteins in other eukaryotes.

Cytoplasmic Hsp90 chaperones are essential proteins required for growth and development of many protozoa (8-10, 17, 18). Lethality of Hsp90 gene knock outs in many eukaryotes (19, 20) and sensitivity to its inhibitor Geldanamycin (GA) in several protozoa including *Eimeria* sp., *L. donovani*, *P. falciparum*, *Toxoplasma gondii*, as well as *Trypanosoma* sp. (9, 10, 21-22) has been demonstrated. We treated *G. lamblia* cells with the GA derivative 17-AAG in vitro and found a dose dependent inhibition of growth. The IC50 value was calculated at 711nM, which is in accordance with that in other organisms.

A long mRNA spanning both the HspN and HspC sequences is present in the transcriptome - Identification of a large protein with the anti-Hsp90 antiserum indicated the presence of a full-sized Hsp90 in *G. lamblia*. Based on the available data we predicted the presence of an mRNA combining the HspN and HspC fragments in the *G. lamblia* transcriptome. To test this hypothesis we used RT-PCR to amplify a predicted full Hsp90 mRNA as well as delineate a possible junction between the two Hsp90 fragments from total RNA. The external primers covering the start of the HspN and the end of the HspC coding regions as shown in GiardiaDB yielded a PCR product of ~2.1 kb (Figure 3A). A smaller, 635 nt fragment was amplified with internal primers HspN-int and HspC-int. This product was not obtained if genomic DNA was used as a template in the reaction as a control (Figure 3B). Sequencing of both the products revealed a contiguous coding region of 2112 nucleotides starting with the predicted ATG codon of ORF 98054 (HspN) and ending with the predicted stop codon of ORF 13864 (HspC). We reasoned that the sequenced mRNA was a splicing product combining the HspN fragment somewhere upstream of the

predicted stop codon with the HspC fragment. Indeed, we could determine a transition from one into the other at 1037 nt downstream of the predicted start codon inside the ORF 98054 coding region (Figure 3C). The opposite splicing site in the HspC fragment localized 99 nt upstream of the predicted ORF 13864, however. Thus, while the splicing reaction discarded 30 codons of ORF 98054, 33 codons were added from the region upstream of the predicted ORF 13864. The combined mRNA codes for a giardial Hsp90 of 704 amino acids whose closest homologs are the Hsp90 proteins in the related species *Spironucleus barkhanus* (696 amino acids, 72% identity) and *Retortamonas* sp. (602 aa, 68% identity). We ruled out the possibility of a putative large pre mRNA which on *cis*-splicing could result in mature Hsp90 mRNA. We have attempted amplification of the region encompassing coding region of HspC and 197 bp downstream of HspC in the non-coding region (arrows in red in Figure 3D) using a PCR-based approach. Using the primer set described in experimental procedures we used either genomic DNA or cDNA as template. As shown in Figure 3D, on using cDNA as template, there was no amplicon obtained on RT-PCR using the primer set corresponding to HspC ORF and downstream of it (lane 1). On the other hand, specific amplification was observed when genomic DNA was used as a template with the same set of primers (lane 2, Figure 3D). In case of a *cis*-splicing event we would have observed a 1172 bp long amplicon on RT-PCR. Absence of a specific product on RT-PCR, confirmed the absence of putative 779 kb long transcript in the RNA pool and thus ruling out *cis*-splicing as a mechanism of RNA processing.

Based on these findings we postulate that *G. lamblia* uses splicing of two separate mRNAs in *trans* as a means to correct the genetic split of Hsp90 and to generate a mature mRNA coding for the ~80 kDa

Hsp90 protein identified by Western blot. This is supported by the presence of canonical GU and AG dinucleotides at the “intron” borders (Figure 3C), suggesting that the two pre mRNAs are substrates for the conventional *cis*-splicing machinery. This leaves the question of how they could be spatially linked to allow a splicing reaction. Interestingly, we identified an almost perfectly matching, complementary region 6 nt downstream of the splicing junction in the HspN sequence, and 35 nt upstream of the splice site in the HspC sequence. As shown in Figure 3E, base pairing of the 26 nt regions of the pre mRNAs would bring the splicing sites into close proximity.

DISCUSSION

Hsp90 is a highly conserved molecular chaperone that assists protein folding and participates in regulation of cell cycle as well as in signal transduction pathways in eukaryotes. The list of clients regulated by Hsp90 is growing as are its roles in different biological processes. With the exception of *Giardia* homolog, Hsp90 is coded by a single gene in all biological systems examined so far and has been shown to possess variable number of introns (23). It is constitutively expressed at high levels under all stages of cell cycle, development and is capable of further induction in response to stress (24, 25).

Examination of *G. lamblia* genomic sequence revealed GIHsp90 to be represented as two separate genes, interrupted by a 770 kb stretch. On multiple sequence alignment of GIHsp90 with canonical Hsp90s we observed the two genes, namely HspN and HspC, to align with N-termini and C-termini of canonical Hsp90 gene sequences. To rule out the possibility that the split Hsp90 gene was a result of erroneously aligned genome fragments, we re-sequenced the whole genome of *G. lamblia* and confirmed that indeed the N and C termini of GIHsp90 are coded by two independent

genes separated by 770 kb intervening sequence. This sole example of a split Hsp90 gene intrigued us to further study Hsp90 from *G. lamblia*.

Are these fragments independently expressed or is there a post-transcriptional event that would result in the generation of a full length Hsp90 gene product? Using primers specific to HspN and HspC and RT-PCR approach we were able to detect an amplicon corresponding to the full length GIHsp90 message. Sequencing of this amplicon confirmed the presence of a fused message arising from genes corresponding to the N- and C-terminal fragments of GHsp90. The fused message was capable of giving rise to a full length GIHsp90 protein bearing all the hallmarks of a canonical Hsp90. We investigated the presence of a full length GIHsp90 using western blotting as well as mass spectrometry approach. Indeed western blotting analysis using antibodies specific to GIHsp90 revealed the presence of a full length GIHsp90 gene product migrating the expected size of about 80 kDa. Partial sequencing of the corresponding protein band from total lysate using mass spectrometry confirmed the presence of peptides corresponding to GIHsp90. Our results proved beyond doubt the presence of a single, full length GIHsp90 protein arising out of HspN and HspC genes interspersed by 770 kb sequence.

G. lamblia is tetraploid with two equivalent diploid nuclei and proliferates asexually in vertebrate hosts. The 12 Mbp genome is very compact with only four genes containing a single intron. A genome rearrangement placing >770 kb of sequence inside the Hsp90 ORF would have been compensated by the remaining intact alleles. However, the current split gene structure in *Giardia* which apparently occurred post speciation shows that this disruption has been genetically fixed. Since Hsp90 function is essential, as shown by inhibition of *Giardia* growth after treatment with Geldanamycin

derivative, this required an alternative mechanism to produce a full-length protein. Absence of the putative 779 kb long pre mRNA transcript formally ruled out the possibility of *cis*-splicing as a potential mechanism to generate a full length mature mRNA. Based on examination of mRNA and genomic sequences a *trans*-splicing reaction directed by canonical dinucleotides GU-AG is the only plausible explanation for the production of a mature mRNA containing both segments of Hsp90. *Giardia* possesses the necessary machinery to remove introns in four genes (26), but RNA *trans*-splicing has not been documented. In trypanosomatids and nematodes this post transcriptional reaction adds capped spliced leader sequences to the 5'-terminal regions of pre-RNAs to generate mature mRNAs from polycistronic precursors (27). There is no example in the literature for the joining of two separate coding RNAs by *trans*-splicing, although synthetic adenovirus encoded mRNAs are spliced in *trans* to allow production of large proteins for gene therapy (28). Although the exact mechanism of the splicing reaction in *Giardia* remains to be determined, the presence of complementary sequence adjacent to the splice sites suggests a positioning function for the assembly of a splicing complex.

Hsp90 is one of the most highly conserved proteins in biological kingdom with limited gene variation within and across species. In addition to its conserved primary structure, its oligomeric structure as well as higher order structure as a multi-chaperone complex is well conserved. *G. lamblia* genome also shows presence of all the members of Hsp90 co-chaperone team described in other systems. It is unclear what advantage has led to the selection of a split Hsp90 gene. The occurrence of a post transcriptional repair suggests that one or both fragments may lead to production of partial proteins which exert distinct functions. Alternatively, repair of split genes may be a means to increase complexity in a virtually intron-less eukaryote. Indeed, the presence of fragmented as well as full-length dynein heavy chain genes in the *Giardia* genome may support this scenario. The documented post transcriptional repair of giardial Hsp90 may be the first example of an ancient mechanism for rescuing disruption of essential genes.

REFERENCES

1. Carlton, J. M., Hirt, R. P., Silva, J. C., Delcher, A. L., Schatz, M., Zhao, Q., Wortman, J. R., Bidwell, S. L., Alsmark, U. C., Besteiro, S., Sicheritz-Ponten, T., Noel, C. J., Dacks, J. B., Foster, P. G., Simillion, C., Van de Peer, Y., Miranda-Saavedra, D., Barton, G. J., Westrop, G. D., Muller, S., Dessi, D., Fiori, P. L., Ren, Q., Paulsen, I., Zhang, H., Bastida-Corcuera, F. D., Simoes-Barbosa, A., Brown, M. T., Hayes, R. D., Mukherjee, M., Okumura, C. Y., Schneider, R., Smith, A. J., Vanacova, S., Villalvazo, M., Haas, B. J., Perte, M., Feldblyum, T. V., Utterback, T. R., Shu, C. L., Osoegawa, K., de Jong, P. J., Hrdy, I., Horvathova, L., Zubacova, Z., Dolezal, P., Malik, S. B., Logsdon, J. M., Jr., Henze, K., Gupta, A., Wang, C. C., Dunne, R. L., Upcroft, J. A., Upcroft, P., White, O., Salzberg, S. L., Tang, P., Chiu, C. H., Lee, Y.

- S., Embley, T. M., Coombs, G. H., Mottram, J. C., Tachezy, J., Fraser-Liggett, C. M., and Johnson, P. J. (2007) *Science* **315**(5809), 207-212
2. El-Sayed, N. M., Myler, P. J., Bartholomeu, D. C., Nilsson, D., Aggarwal, G., Tran, A. N., Ghedin, E., Worthey, E. A., Delcher, A. L., Blandin, G., Westenberger, S. J., Caler, E., Cerqueira, G. C., Branche, C., Haas, B., Anupama, A., Arner, E., Aslund, L., Attipoe, P., Bontempi, E., Bringaud, F., Burton, P., Cadag, E., Campbell, D. A., Carrington, M., Crabtree, J., Darban, H., da Silveira, J. F., de Jong, P., Edwards, K., Englund, P. T., Fazelina, G., Feldblyum, T., Ferella, M., Frasch, A. C., Gull, K., Horn, D., Hou, L., Huang, Y., Kindlund, E., Klingbeil, M., Kluge, S., Koo, H., Lacerda, D., Levin, M. J., Lorenzi, H., Louie, T., Machado, C. R., McCulloch, R., McKenna, A., Mizuno, Y., Mottram, J. C., Nelson, S., Ochaya, S., Osoegawa, K., Pai, G., Parsons, M., Pentony, M., Pettersson, U., Pop, M., Ramirez, J. L., Rinta, J., Robertson, L., Salzberg, S. L., Sanchez, D. O., Seyler, A., Sharma, R., Shetty, J., Simpson, A. J., Sisk, E., Tammi, M. T., Tarleton, R., Teixeira, S., Van Aken, S., Vogt, C., Ward, P. N., Wickstead, B., Wortman, J., White, O., Fraser, C. M., Stuart, K. D., and Andersson, B. (2005) *Science* **309**(5733), 409-415
 3. Ivens, A. C., Peacock, C. S., Worthey, E. A., Murphy, L., Aggarwal, G., Berriman, M., Sisk, E., Rajandream, M. A., Adlem, E., Aert, R., Anupama, A., Apostolou, Z., Attipoe, P., Bason, N., Bauser, C., Beck, A., Beverley, S. M., Bianchetti, G., Borzym, K., Bothe, G., Bruschi, C. V., Collins, M., Cadag, E., Ciarloni, L., Clayton, C., Coulson, R. M., Cronin, A., Cruz, A. K., Davies, R. M., De Gaudenzi, J., Dobson, D. E., Duesterhoeft, A., Fazelina, G., Fosker, N., Frasch, A. C., Fraser, A., Fuchs, M., Gabel, C., Goble, A., Goffeau, A., Harris, D., Hertz-Fowler, C., Hilbert, H., Horn, D., Huang, Y., Klages, S., Knights, A., Kube, M., Larke, N., Litvin, L., Lord, A., Louie, T., Marra, M., Masuy, D., Matthews, K., Michaeli, S., Mottram, J. C., Muller-Auer, S., Munden, H., Nelson, S., Norbertczak, H., Oliver, K., O'Neil, S., Pentony, M., Pohl, T. M., Price, C., Purnelle, B., Quail, M. A., Rabinowitsch, E., Reinhardt, R., Rieger, M., Rinta, J., Robben, J., Robertson, L., Ruiz, J. C., Rutter, S., Saunders, D., Schafer, M., Schein, J., Schwartz, D. C., Seeger, K., Seyler, A., Sharp, S., Shin, H., Sivam, D., Squares, R., Squares, S., Tosato, V., Vogt, C., Volckaert, G., Wambutt, R., Warren, T., Wedler, H., Woodward, J., Zhou, S., Zimmermann, W., Smith, D. F., Blackwell, J. M., Stuart, K. D., Barrell, B., and Myler, P. J. (2005) *Science* **309**(5733), 436-442
 4. Adam, R. D. (2000) *Int J Parasitol* **30**(4), 475-484
 5. Morrison, H. G., McArthur, A. G., Gillin, F. D., Aley, S. B., Adam, R. D., Olsen, G. J., Best, A. A., Cande, W. Z., Chen, F., Cipriano, M. J., Davids, B. J., Dawson, S. C., Elmendorf, H. G., Hehl, A. B., Holder, M. E., Huse, S. M., Kim, U. U., Lasek-Nesselquist, E., Manning, G., Nigam, A., Nixon, J. E., Palm, D., Passamaneck, N. E., Prabhu, A., Reich, C. I., Reiner, D. S., Samuelson, J., Svard, S. G., and Sogin, M. L. (2007) *Science* **317**(5846), 1921-1926
 6. Simpson, A. G., and Roger, A. J. (2002) *Curr Biol* **12**(20), R691-693
 7. Neckers, L., and Tatu, U. (2008) *Cell Host Microbe* **4**(6), 519-527
 8. Sawarkar, R., Roy, N., Rao, S., Raman, S., Venketesh, S., Suguna, K., and Tatu, U. (2008) *J Mol Biol* **383**(1), 24-35
 9. Wiesgigl, M., and Clos, J. (2001) *Mol Biol Cell* **12**(11), 3307-3316
 10. Pallavi, R., Roy, N., Nageshan, R. K., Talukdar, P., Pavithra, S. R., Reddy, R., Venketesh, S., Goel, R. K., Gupta, A. K., Singh, R. K., Yadav, S. C., and Tatu, U. *J Biol Chem*
 11. Rutherford, S. L., and Lindquist, S. (1998) *Nature* **396**(6709), 336-342

12. Queitsch, C., Sangster, T. A., and Lindquist, S. (2002) *Nature* **417**(6889), 618-624
13. Cowen, L. E., and Lindquist, S. (2005) *Science* **309**(5744), 2185-2189
14. Diamond LS, Harlow DR, Cunnick CC (1978) *Trans R Soc Trop Med Hyg* **72**(4): 431-432.
15. Yee, J., and Dennis, P. P. (1992) *J Biol Chem* **267**(11), 7539-7544
16. Harris, J. C., Plummer, S., Turner, M. P., and Lloyd, D. (2000) *Microbiology* **146 Pt 12**, 3119-3127
17. Peroval, M., Pery, P., and Labbe, M. (2006) *Int J Parasitol* **36**(10-11), 1205-1215
18. Ahn, H. J., Kim, S., and Nam, H. W. (2003) *Biochem Biophys Res Commun* **311**(3), 654-659
19. Devaney E, O'Neill K, Harnett W, Whitesell L, Kinnaird JH. (2005) *Int J Parasitol.* **35**(6):627-36.
20. Borkovich KA, Farrelly FW, Finkelstein DB, Taulien J, Lindquist S (1989) *Mol Cell Biol.* **9**(9): 3919-3930.
21. Echeverria, P. C., Matrajt, M., Harb, O. S., Zappia, M. P., Costas, M. A., Roos, D. S., Dubremetz, J. F., and Angel, S. O. (2005) *J Mol Biol* **350**(4), 723-734
22. Pavithra, S. R., Banumathy, G., Joy, O., Singh, V., and Tatu, U. (2004) *J Biol Chem* **279**(45), 46692-46699
23. Chen, B., Piel, W. H., Gui, L., Bruford, E., and Monteiro, A. (2005) *Genomics* **86**(6), 627-637
24. Pan, F., Zarate, J. M., Tremblay, G. C., and Bradley, T. M. (2000) *J Exp Zool* **287**(3), 199-212
25. Chen, B., Kayukawa, T., Monteiro, A., and Ishikawa, Y. (2005) *Insect Mol Biol* **14**(6), 697-702
26. Nixon JE, Wang A, Morrison HG, McArthur AG, Sogin ML, Loftus BJ, Samuelson J. (2002) *Proc Natl Acad Sci U S A* **99**(6): 3701-3705.
27. Hannon GJ, Maroney PA, Yu YT, Hannon GE, Nilsen TW (1992) *Science* **258**(5089): 1775-1780.
28. Nakai H, Storm TA, Kay MA (2000) *Nat Biotechnol* **18**(5): 527-532.

FOOTNOTES

Authors thank Ms. Shatakshi Ranade for her valuable inputs in the Hsp90 inhibitor sensitivity experiment. This work was supported by Department of Science and Technology, New Delhi, Department of Biotechnology, New Delhi, India, and MALSIG. R.K.N and N.R acknowledge Junior Research Fellowship from the Council of Scientific and Industrial Research and United Grants Commission, New Delhi, India. The work in the laboratory of A.B.H. is supported by the Swiss National Science Foundation (grant #31003A125389).

FIGURE LEGENDS

FIGURE 1: Hsp90 is arranged as a split gene in *G. lamblia* genome. A, Representation of gene position of GIHsp90 as reported previously by Morrison *et al.*, Science, 2007. B, Graphical representation of fused cDNA of GIHsp90, with HspN and HspC fused in head to tail pattern. C, PCR amplification using specific primers against the flanking region of HspN and HspC (200 bp upstream sense primer and downstream antisense primer) gives a specific 1.5 kb (HspN) and 1.4 kb (HspC) long product respectively, confirming the genomic position and organization of Hsp90 split gene.

FIGURE 2: A full length cytosolic Hsp90 in *G. lamblia*. *A*, GIHsp90 is expressed as a full length, 80 kDa protein. Lane 1 – Total protein profile from Trophozoite stage of *G. lamblia* cells fractionated on 10% SDS PAGE stained with Coomassie Brilliant Blue. Lane 2. Western blot analysis of total protein probed with anti GIHsp90 antibody showing a specific band corresponding to 80 kDa. *B*, MS/MS spectrum of signature peptides from HspN and HspC from the 80 kDa band. *C*, GIHsp90 is essential for growth of parasites *in vitro*. The graph represents percentage survival of parasites against increasing concentrations of 17AAG. The IC50 value of 17AAG was determined to be 711nM *in vitro*. *D*, Indirect immunofluorescence analysis using anti GIHsp90 antibody shows a diffused cytosolic localization of GIHsp90 in the trophozoite stage of the parasite. *E*, HspN and HspC peptides identified by MS/MS analysis are highlighted on the respective protein sequences.

FIGURE 3: Splicing junction of GIHsp90. *A*, Full length GIHsp90 amplified from cDNA, using sense primer from HspN and antisense primer against HspC. *B*, A PCR product spanning the splice site of the mature Hsp90 mRNA could be amplified with internal primers from cDNA (lane 1) but not from genomic DNA (lane 2). *C*, Graphical depiction of the complementary 26-mer motifs (red boxes) on the two Hsp90 pre-mRNAs. Base pairing would bring the cleavage sites (asterisks) into close approximation. The respective sequences and orientations are shown below. Dotted lines indicate ORF boundaries which do not correspond to those predicted in GiardiaDB. *D*, Representation of primers on the 777 kb fragment showing the annealing sites at HspC ORF and downstream flanking region. Arrows in red represents primer annealing sites at HspC ORF and downstream flanking region. Arrows in black represents direction of transcription of HspC ORF. Amplification using forward primer annealing within HspC ORF and reverse primer against flanking region of HspC (197 bp downstream) could not amplify gene product from RNA (lane 1) whereas specific product was obtained on using genomic DNA as template (lane 2). Middle domain served as a positive control (lane 3) for RNA. *E*, Alignment showing the “exon” borders (top and bottom lines) and the transition sequence in the mature GIHsp90 mRNA (middle line) determined by sequencing of cDNAs. The sequence segments coding for predicted Hsp90 fragments in GiardiaDB are underlined. The GU and AG dinucleotides at the “intron” borders are boxed. Discarded sequences are shown in italic letters. The transition point in the sequence of the spliced mRNA is depicted as an angled red line. The numbers in brackets indicate the number of nucleotides. Predicted start and stop codons are in boldfaced letters.

Figure 1

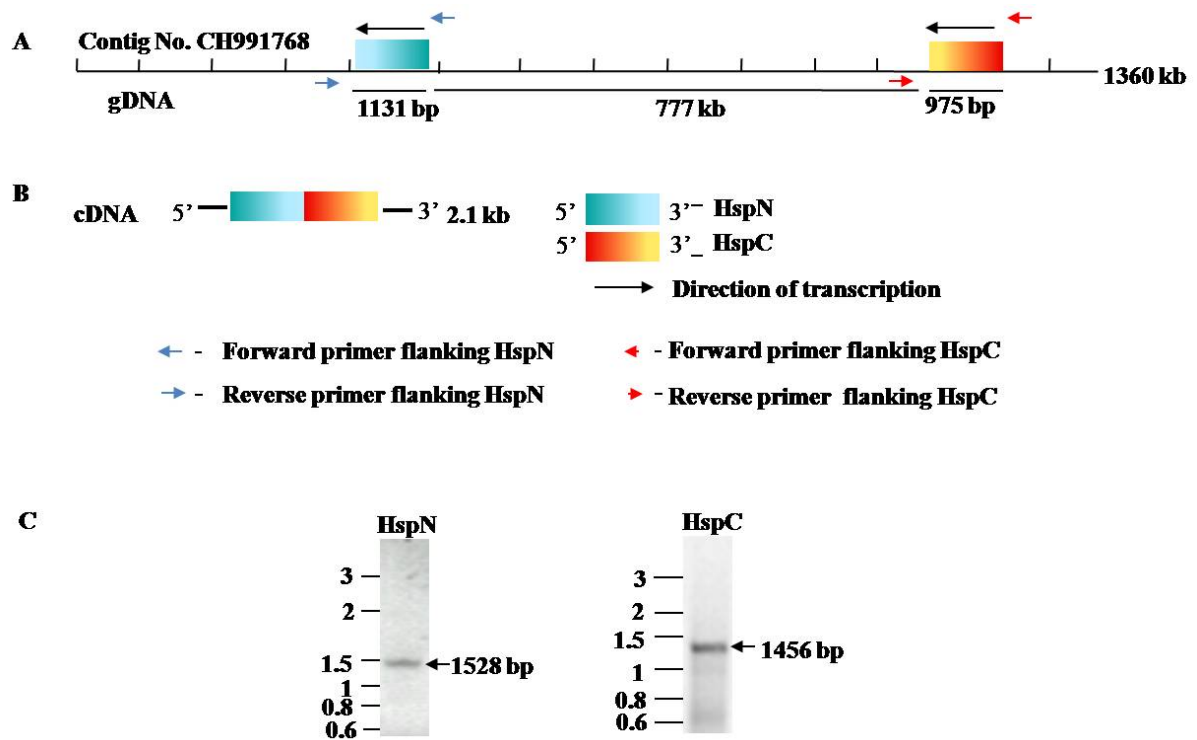


Figure 2

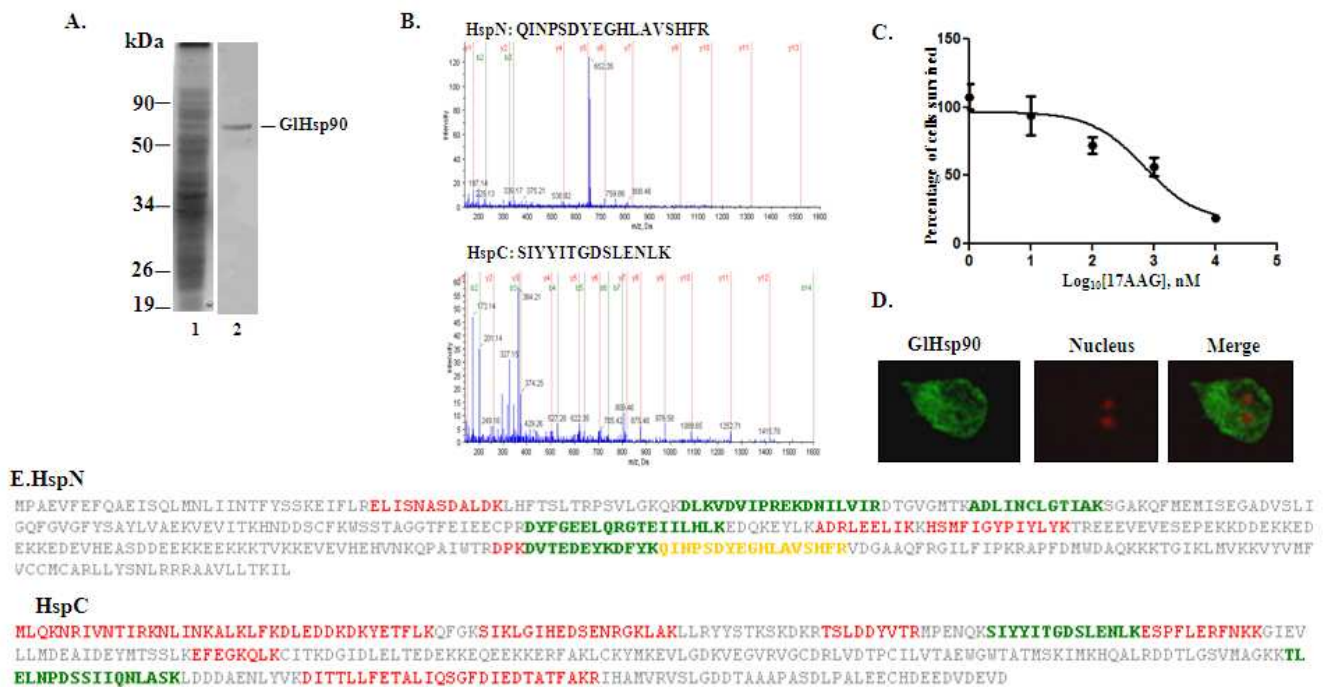


Figure 3

